

博士論文（要約）

ディープラーニングを用いた胸部 X 線の読影による
成人心房中隔欠損症のスクリーニングについての検討

松岡 良

論文の内容の要旨

論文題目 ディープラーニングを用いた胸部 X 線の読影による成人心房中隔欠損症のスクリーニングについての検討

氏名 松岡 良

(序文)

心房中隔欠損症(atrial septal defect, ASD)は先天性心疾患の中で頻度の高い疾患であり、特に成人先天性心疾患の中では最も頻度が高い。他の先天性心疾患を合併しない単独の ASD の多くは小児期から若年期にかけて無症状であるが、無治療のままで経過すると年齢とともに徐々に心不全や不整脈が出現し、また脳梗塞などの血栓症を合併することもあり、最終的には生存期間の低下をもたらす。成人患者においても ASD 閉鎖術は予後を有意に改善することが示されており、成人になってから診断された ASD でも閉鎖術の適応を検討すべきとされている。しかし、25 歳未満で閉鎖術を施行された患者は健常者と同等の予後が期待できる一方で、25 歳以上で施行された患者は健常者と比較した場合の予後が有意に悪く、閉鎖術施行時の年齢が高くなるほど長期的に心不全や心房細動、脳梗塞を発症する頻度が有意に多くなるという報告もあり、予後を改善するためには早期診断と早期の閉鎖術施行が望ましい。

しかし、ASD は多くの症例は小児期から若年期にかけて多くが無症状であり、自発的な病院受診は期待できない。また、健康診断で施行される聴診、心電図、胸部 X 線検査は単独では感度が高いとはいえない。それでも多くの症例は学校検診などで小児期のうちに診断されるが、成人になってから健康診断などで初めて ASD が発見される症例や高齢になって心不全増悪や脳梗塞を発症して初めて診断される症例も稀ではなく、このような見逃し例をどのように防ぐかが課題である。

こうした診断の補助として期待されているのが人工知能 (AI) 技術である。最近の AI 技術の発展は医学分野においても著しく、特に画像診断においてはディープラーニングを用いた手法が医師に匹敵する精度での診断が可能であったとする報告が多数存在している。ASD は胸部 X 線で特徴的な所見をとるため、AI 技術を用いた読影によるスクリーニングが可能ではないかと仮説を立てた。

今回の研究では、ディープラーニングの手法で胸部 X 線画像から ASD 患者と非 ASD 患者を見分けるシステムを作成し、実際の患者データで循環器内科医と同等の精度での診断が可能であることを確認した。このシステムを日常診療や健康診断に適用することで、これまで未診断であった成人 ASD を早期に診断し、ひいては早期治療と予後改善に繋げることを目標としている。

(方法)

ASD 群として、2005 年 4 月から 2020 年 6 月に東京大学医学部附属病院循環器内科を受診した受診時点で 15 歳以上の ASD と確定診断済みの患者を対象とした。その中で ASD 閉鎖前の胸部 X 線画像が残っていない患者、ペースメーカー挿入前の胸部 X 線画像が残っていない患者、当院での検査時点でアイゼンメンジャー症候群を来している患者は除外した。対象患者から電子カルテでアクセスできる範囲で ASD 閉鎖術を行う前の画像を患者あたり複数枚の画像を許容して収集した。非 ASD 群として、①上記期間内に同科に虚血性心疾患（疑い）に対して冠動脈造影検査または経皮的冠動脈形成術を受ける目的で入院した、②入院時点で 15 歳以上、③少なくとも心臓超音波検査で ASD が指摘されていない、④心臓手術を受けたことがない、⑤胸部 X 線が急性心不全の所見を呈していない、という条件を満たす患者を候補とし、ASD 群と平均年齢や性別比が近くなるように手動で選択した。対象患者から ASD 群と同様に胸部 X 線画像を収集した。ASD 群と非 ASD 群の画像を全て合わせたものを ASD データセットと定義した。また、ASD 群の患者についてはわかる範囲での心臓超音波検査や右心カテーテル検査の結果を収集した。

プログラミング言語 Python によるプログラミングで画像分類を行うディープラーニングモデルを作成した。これには転移学習の手法を用いたが、これは別のデータセットで学習を行ったベースモデルに目的のデータセットを再学習させ、目的のデータセットに使えるように流用する手法である。ベースモデルとして、以前の画像分類コンペティションで優秀な成績を収めたモデルである VGG16 に ImageNet という自然画像のデータセットを学習させたモデル (IN-VGG16) と、VGG16 の構造を一部変更したものに胸部 X 線画像の心陰影拡大の有無を学習させたモデル (Cardiomegaly-modified VGG16, CM-mVGG16) の 2 種類を作成した。

それぞれのモデルに ASD データセットの一部（トレーニングセット）を学習させ、IN-ASD-VGG16、CM-ASD-mVGG16 モデルとした。これらに ASD データセットの残り（テストセット）を読影させ、ROC 曲線と曲面下面積（AUC）、感度、特異度、正答率を調べた。また、同じデータを循環器内科医 3 名に読影させ、平均の診断精度を評価してモデルと比較した。モデルが画像のどの範囲を診断基準としているかを gradient-weighted class activation mapping (Grad-CAM) の手法で可視化した。モデルの診断と患者背景を比較し、相関を調べた。

（結果）

対象期間に東京大学医学部附属病院循環器内科を受診した 15 歳以上の ASD 患者は 99 名であった。このうち、ASD 閉鎖術施行前の胸部 X 線画像がない患者（ $n = 20$ ）、ペースメーカー挿入術前の胸部 X 線画像がない患者（ $n = 2$ ）、当院での検査時点でアイゼンメンジャー症候群を呈していた患者（ $n = 3$ ）は除外し、残りの 74 名の患者から胸部 X 線画像を収集すると、173 枚であった。一方非 ASD 群として前述した通りに患者を選択し、患者 100 名から計 174 枚の胸部 X 線画像を取得した。

IN-VGG16 および CM-mVGG16 の 2 つの学習済みモデルを作成し、それぞれに同じ ASD トレーニングセットを学習させた。その後、それぞれのモデルに同じ ASD テストセットを入力し、診断精度を評価した。IN-ASD-VGG16 モデルでは、AUC 0.94、感度 77%、特異度 97%、正答率 87%であった。CM-ASD-mVGG16 モデルでは、AUC 0.96、感度 100%、特異度 85%、正答率 93%であった。同じテストセットを 3 人の循環器内科医が読影した時の平均は、AUC 0.90、感度 89%、特異度 79%、正答率 84%であった。それぞれのモデルと循環器内科医の AUC の間に統計学的に有意な差は認めなかった。

Grad-CAM で 2 つのモデルが入力画像のどの部分に注目しているかを評価した。IN-ASD-VGG16 モデルは左肺門部から心陰影左側にかけての範囲に注目する傾向があった。一方、CM-ASD-mVGG16 モデルは右肺門部から心陰影右側にかけての範囲を見て判断を行う傾向があった。この違いがそれぞれのモデルの感度・特異度の違いに寄与しているようであった。

入力画像に対する各モデルの出力と、入力画像に対応する心臓超音波検査および右心カテーテル検査の検査値の相関を検討した。心臓超音波検査では IN-ASD-VGG16 と推定右房圧、推定右室圧に有意な相関を認めた。右心カテーテル検査では統計学的には有意な相関は認めなかったが、IN-ASD-VGG16 は平均肺動脈圧 ($R = 0.46$)、CM-ASD-mVGG16 は平均右房圧 ($R = 0.47$) と最も高い相関係数を有していた。

(考察)

胸部 X 線画像をディープラーニングモデルで読影したという研究は多いが、成人 ASD の診断にディープラーニングモデルを使用したのは本研究が初である。

実臨床の面で言えば、今回作成したディープラーニングモデルは胸部 X 線画像のみから高精度に ASD を診断することが可能であり、これまで見逃されていた ASD を早期に診断できる可能性がある。特に本邦では健康診断や人間ドック、がん検診などへのアクセスが容易であり他国と比較しても無症状の時点で胸部 X 線検査を受ける機会がより多いと考えられ、今回の手法を用いたスクリーニングはより有効であろう。また、読影医の負担を軽減する効果も期待できる。

Grad-CAM ではモデルによって画像の注目範囲が異なり、感度の比較的高い CM-ASD-mVGG16 は右房～右肺門部付近を、特異度の高い IN-ASD-VGG16 は左肺門部付近を主な判断基準として見ていることが示された。患者背景と比較しても CM-ASD-mVGG16 の診断は平均右房圧と、IN-ASD-VGG16 の診断は平均肺動脈圧と高い相関係数を持っており、病態と画像が一致していることが伺える。今後、画像から ASD の重症度を評価することも可能になるかもしれない。

(結論)

実際の成人 ASD 患者の胸部 X 線画像を学習させることで、胸部 X 線画像から ASD を診

断するディープラーニングモデルを作成した。実際の患者の画像で診断精度を確認し、感度、特異度ともに一般循環器内科医と同等であることを確認した。このモデルを日常診療や健康診断の胸部 X 線検査に適用することで、これまで未診断であった ASD を早期に発見し、予後を改善できる可能性がある。実臨床への応用に向け、今後多施設のデータを含むより大きなデータセットを用いた研究を進めていく。